

DOCUMENT RESUME

ED 202 410

HE 013 932

AUTHOR Aubrecht, Judith D.
 TITLE Are Student Ratings of Teacher Effectiveness Valid?
 IDEA Paper No. 2.
 INSTITUTION Kansas State Univ., Manhattan. Center for Faculty
 Evaluation and Development in Higher Education.
 PUB DATE Nov 79
 NOTE 7p.; Not available in paper copy due to small
 print.
 AVAILABLE FROM Kansas State University, Center for Faculty
 Evaluation and Development, 1627 Anderson Ave., Box
 3000, Manhattan, KS 66502.
 EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
 DESCRIPTORS *College Faculty; College Students; Evaluation
 Criteria; Grading; Higher Education; Instructional
 Improvement; Reliability; Student Attitudes; *Student
 Evaluation of Teacher Performance; Teacher Behavior;
 Teacher Characteristics; *Teacher Effectiveness;
 *Teacher Evaluation; Teacher Improvement;
 *Validity

ABSTRACT

Issues pertaining to the validity of student ratings of teaching effectiveness are considered, and research results are evaluated. The following extraneous variables that affect student ratings but are not under the control or influence of the teacher are examined: class size, initial student motivation, academic field, teacher characteristics, and situational leniency. If effective teaching is defined by student progress on teacher defined goals, some measure of that progress is needed. The following measures are assessed: grades, achievement tests, and student self-report. Based on the research, it is suggested that student ratings are adequately reliable and relevant measures to be included in teacher evaluation procedures. Evidence is cited to indicate that student ratings are strongly related to the ratings of teacher effectiveness by administrators, peers, and the teachers themselves. It was also found that students can make fine distinctions along a number of dimensions of teacher effectiveness. It was also found that student ratings by themselves are not very helpful to teachers seeking to improve instruction. However, some student rating systems (notably IDEA) provide interpretative assistance that can enhance the usefulness of student ratings for diagnosis. It is also possible that instructional development specialists can provide help to teachers in translating the information from student ratings into appropriate changes in teacher behavior. A bibliography is included. (SW)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

*Center for Faculty
& Evaluation of the
in H.S. Kansas State*

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

O-
M-
N-
IS
E-

center for
FACULTY
EVALUATION &
DEVELOPMENT

IN HIGHER EDUCATION
KANSAS STATE UNIVERSITY

IDEA PAPER No. 2

November, 1979

Are Student Ratings of Teacher Effectiveness Valid?

Judith D. Aubrecht
Kansas State University

What is effective teaching?

Effective is obviously not a universal or absolute term. We will have to ask, "Effective for what purpose?" Effective teaching could be defined in terms of the teacher's goals, the department's goals, the goals of the particular profession, the institution's goals, or the student's goals. All of these would be legitimate educational ends. We will focus on the goals of the teacher for a particular course - accepting the notion that the teacher is the best mediator of the goals of all of the above sources.

A working definition: Teacher effectiveness is the amount of progress students make on teacher defined goals. (This definition is the heart of the Center's IDEA* system.)

Obviously, we could argue with this definition - and will, briefly.

Objection I: The teacher may in fact not pick worthwhile goals. Using our definition if the students make great progress on trivial goals, we still have to agree that the teaching was "effective."

Objection II: Students may learn without the teacher's assistance or even in spite of the teacher. We are assuming that teachers do generally have and use the ability to assist student learning.

Objection III: The teacher could promote student learning in ways we disapprove - indoctrination, coercion, etc. - and still be rated "effective" using our definition. We are assuming the teacher's good intent and good sense.

Flawed though our definition may be, or at least open to the above objections, it has certain advantages.

Advantage I: We are not assuming any single "ideal" teacher model. There are many effective ways of teaching. They vary with the course content, the teacher, the students, etc.

Advantage II: Our definition does tie teaching to its intended outcome - student learning. It keeps the goals of teaching clearly in mind. It may be possible to have teaching without learning. It is not possible to have effective teaching without learning.

What is validity?

Validity is the extent to which ratings measure what they are intended to measure. No set of scores is "valid" in any absolute sense. We have to ask, "How valid for what purpose?" Basically student ratings are used for two purposes; evaluating teachers (for promotion, tenure, merit pay, etc.) and providing diagnostic information to teachers (for improvement of instruction). A rating system that is relatively valid for one of these purposes will not necessarily be valid for the other (just as a valid test of skill in manipulating fractions is not necessarily a valid test for knowledge of French vocabulary). If students give high ratings to teachers in courses in which they learn a great deal (and low ratings in courses in which they do not) then the ratings have validity for our first purpose (evaluation). If students' ratings can be used to suggest changes in teacher behavior which are likely to improve instruction, then the ratings have validity for our second purpose (diagnosis and improvement of instruction).

In order for student ratings to be valid, they must be both reliable (they measure consistently whatever they do measure) and relevant (what they measure is what they ought to measure for the purposes they serve). The ratings could be reliable (consistent, replicable) but turn out to measure primarily the degree to which the student was motivated to learn the material before the course began, rather than the influence of the teacher on learning. In this case we would not want to draw conclusions about teacher effectiveness from student ratings.

No one has found perfectly reliable and relevant (and therefore perfectly valid) measures of teacher effectiveness. There are no micrometers in this area. Yardsticks would be great but we will probably be happy to settle for the distance from elbow to tip of middle finger.

If effective teaching is defined by the progress students make on teacher defined goals, how can we test the validity of student ratings of teacher effectiveness? There are several questions we need to ask, and there is a growing body of literature on which to draw for some tentative answers.

ED 202410

HE 013932

I. For both evaluative and diagnostic purposes:

A. Are student ratings reliable?

B. Can we identify and control for powerful extraneous variables that affect student ratings but are not under the control or influence of the teacher?

II. For evaluative purposes:

A. How can we measure student progress (e.g., grades, achievement tests, self-reports of perceived learning)?

B. Do student ratings of teacher effectiveness correlate with administrator ratings, peer (colleague) ratings, instructor self-ratings? (This would give us some assurance that students' ideas about teaching are similar to those of professionals.)

C. Do students distinguish degrees and kinds of effective teaching (as opposed to a global "good-bad" distinction)?

III. For diagnostic purposes:

A. Can student ratings directly supply information that will assist the teacher in improvement of instruction?

B. Can student ratings be combined with an empirical research base (correlating student progress and specific teacher activities) to supply information that will assist the teacher in improvement of instruction?

We will draw on the major reviews of the literature in this area and a few selected individual papers in an attempt to throw some light on these questions.

ARE STUDENT RATINGS RELIABLE?

This is probably the best researched and therefore easiest question we have to answer. Reliabilities for student ratings are typically in the .80's and .90's for a class size of twenty or more. (As an example, in the IDEA system the average item reliability is .69 for ten raters, .81 for twenty, .89 for forty.) (See footnote 1 for a more extended discussion.) Student ratings are certainly reliable enough for our purposes.

CAN WE IDENTIFY AND CONTROL FOR POWERFUL EXTRANEIOUS VARIABLES THAT AFFECT STUDENT RATINGS BUT ARE NOT UNDER THE CONTROL OR INFLUENCE OF THE TEACHER?

If student ratings reflected only the content and teaching of a course, we would have no worries about extraneous variables or "bias". However, we know that there are certain powerful correlates of student ratings - initial student motivation and class size being two examples. We can either consider them "biasing" factors or we can consider them legitimate influences on student ratings. In either case, we need to study the influence of these variables to determine how strongly they affect student ratings. (See Feldman, 1977 for a more complete discussion.)

Class Size: Correlations reported in the literature between student ratings and class size generally range from negative .10 to negative .30 (smaller classes receive higher ratings). Several studies have reported a nonlinear relationship with medium size classes receiving lower ratings than either smaller or larger classes. (See Feldman, 1978; Costin, Greenough, and Menges, 1971.) This relationship is moderately powerful and sufficiently well established that the IDEA system controls for class size.

Initial Student Motivation: Doyle and Whitely (1974) reported correlations of .42 and .48 between students' initial liking for the subject and some overall evaluation items in student ratings. IDEA system research indicates that this motivation factor is sufficiently powerful that it is used in combination with class size to establish norm groups for comparison purposes. (See also Clegg, 1979; Howard and Schneck, 1979 for discussions of the impact of student motivation on ratings.)

Academic Field: In studies controlling several other important factors, differences in academic field still generally appear. Teachers in the humanities, fine arts and languages seem to receive higher ratings than teachers in social or physical sciences, mathematics and engineering. (See Feldman, 1978.) The Center is experimenting with the incorporation of academic field into its control procedures for IDEA.

Teacher Characteristics: A few words about teacher characteristics and their relationship to student ratings seem in order although these are hardly "extraneous" variables. (They are certainly under the influence of the teacher.) The relationship between student ratings and either academic rank or research productivity seems to be positive but quite weak. Bendig (1955) found strong correlations between reading comprehension scores (of instructors) and student ratings ($r=.50$). In general, instructors with high ratings seem to differ from those with low ratings on measures of communication ability. (See Kulik and McKeachie, 1975; Kulik and Kulik, 1974.)

Situational Leniency: "Leniency error" apparently exists under certain conditions. That is, students tend to rate instructors higher if the ratings are not anonymous, if the instructor is present while they are completing the forms, or if they are told that the ratings will be used for promotion or salary decisions (rather than course development). When conditions such as the above are known to affect ratings, it is important and not terribly difficult to assure that these conditions are standard across an institution. (Either all instructors should be present during the rating sessions or none should be.) (See Feldman, 1979.)

So far we have established a working definition of teacher effectiveness and certain criteria for answering questions of validity. We have also discussed and arrived at generally positive conclusions to our first two major questions: (1) Are student ratings reliable? and (2) Can we identify and control for powerful extraneous variables that affect student ratings but are not under the control or influence of the teacher? We still need to look at ways of measuring student progress, the relationship between student ratings and colleague, administrator, or self-ratings, and the ability of students to distinguish degrees and kinds of effective teaching. All of these will have a bearing on the validity of student ratings for evaluative purposes. We will also look at the direct diagnostic value of student ratings and the possibility for combining student ratings with empirical data in order to enhance the diagnostic usefulness. These will relate to the validity of student ratings for purposes of instructional improvement.

HOW DO WE MEASURE STUDENT PROGRESS?

If effective teaching is defined by student progress on teacher defined goals, we need some measure of that progress. This measure could be grades, achievement tests, or students' perceptions of their own learning.

Grades: Most of the studies correlating grades and student ratings report correlations from the mid .10's to just below .30. However, these correlations come from pooling data across a number of classrooms, a procedure which may disguise results from individual classes. There is some evidence that grades and ratings may have strong positive correlations in some classrooms, no apparent relationship in others, and even negative correlations in others. (See Feldman, 1976a.) Remmers (1928) and Elliott (1950) suggest that a positive correlation between ratings and grades would be expected if the instructor "teaches to" the better students in the class and negative correlation would be expected if the instructor teaches to the poorer students.

Because of the wide variety in grading practices, the inconsistency of the results in correlating grades and ratings cited above, and the logistical difficulty involved in obtaining student grades and matching them to individual ratings, grades seem to be a poor choice as a measure of student progress on which to judge teacher effectiveness.

Achievement Tests: According to Kulik and Kulik (1974), the median unadjusted correlation reported in the literature between student achievement and student ratings is .23. However, there are some inconsistencies with this measure as there were with grades. Sullivan and Skanes (1974) found that student ratings and achievement were highly related for a group of experienced teachers, but not for a group of inexperienced teachers.

There are great logistical difficulties even in studying the relationship between student ratings and achievement tests. The instructors must be teaching the same course to comparable groups of students and other conditions must also be fairly standard. A few studies have attempted to standardize conditions in this way. In one widely quoted study, eleven teaching assistants taught two recitation sections a week in a large undergraduate calculus class. (Rodin and Rodin, 1973) There was a negative .75 correlation between instructor ratings and class performance on a common examination. There was also considerable methodological criticism of this study. In another (much better designed) study of two large basic calculus courses, Frey (1973) found correlations of .91 and .60 (both positive) between student ratings and student performance on final exams. Solomon, Rosenberg, and Bezdek (1964) also reported positive correlations of .46 between ratings and average factual gains for a group of 24 evening classes. (See Kulik and McKeachie, 1975; Kulik and Kulik, 1974.)

There is more consistency in the research on the relationship between achievement tests and student ratings than we found on the relationship between grades and student ratings. However, if we adopt achievement tests as our measure of student progress we will have to deal with the difficult question of comparing achievement data across fields (or even across courses in the same field).

Student self-report: Lathrop and Richmond (1967) and Lathrop (1968) found positive correlations between student ratings and the extent to which students thought they had achieved course objectives. In general, student self-report of perceived learning and overall ratings of teachers and courses do correlate well (in the .40's and .60's). (See Feldman, 1976a.) There are also studies indicating that student self-reports of progress correlate well with their scores on external measures of achievement. These correlations tend to be .50 or above. (See Frey 1973, 1975, 1976; Marsh 1956; Solomon et al, 1964.)

Student self-reports are well correlated with student ratings and seem to be an adequate measure of actual achievement. They also avoid the difficulties outlined above in the use of achievement data or grades, especially in the comparisons across fields. For these reasons, the IDEA system uses students' perceptions of their own learning as the measure of student progress.

DO STUDENT RATINGS OF TEACHER EFFECTIVENESS CORRELATE WITH ADMINISTRATOR RATINGS, PEER (COLLEAGUE) RATINGS, INSTRUCTOR SELF-RATINGS?

Guthrie (1954) reported a correlation of .63 between student evaluations and colleague ratings on a three-item scale. Maslow and Zimmerman (1956) found a .69 correlation between administrator and colleague ratings and a .47 correlation between administrator and student ratings. Generally, administrator and colleague ratings are nearly interchangeable and each is fairly well correlated with student ratings. However, there is some difficulty in the assumption that colleague or administrator ratings are independent of student ratings. (See Kulik and McKeachie, 1975.) A high correlation between colleague and student ratings (.43 to .69) is found when the colleague ratings are not based on classroom visits (but are probably based mainly on feedback from students). Centra (1975) tried to reduce this dependence on student feedback in a study using colleague classroom visits. This study was set in a new school where there would have been little opportunity for student feedback. Unfortunately, there was little agreement among colleagues ($r=.26$) based on the classroom visits, and therefore little agreement possible between colleague and student ratings ($r=.20$). (See Marsh, Overall and Kesler, 1979.)

Until recently, studies of instructor self-ratings and their relationship to student ratings (or, for that matter, to administrator or colleague ratings) have come up with pretty disappointing results. For example, Blackburn and Clark (1975) found very low correlations between faculty self-evaluations and administrator, student, or peer evaluations. There is a rather crucial issue of "face validity" here. That is, if there is so little agreement between instructor self-ratings and student ratings, it is unlikely that instructors will take student ratings seriously no matter what their merits. However, in the Blackburn and Clark (1975) study just cited (and in other studies of the same question) the faculty self-evaluations were very general impressions of their own teaching effectiveness and were not tied to any particular course. In a very exciting recent study by Marsh, Overall and Kesler (1979) the results were quite different. Fifty-one faculty were asked to identify, from the two courses they had just taught, the course in

which their teaching was more effective and the one in which their teaching was less effective. They were also asked to evaluate their teaching in each course, using the same items that students used to evaluate the same course. This procedure requires some pretty fine distinctions on the part of the teacher. Factor analyses done separately on the faculty forms and the student forms indicated the same underlying dimensions. Validity coefficients for all these factors were significant (median $r = .49$). The mean differences between student and faculty ratings were very small. Apparently when given basically the same rating task, students and teachers are in very good agreement.

DO STUDENTS DISTINGUISH DEGREES AND KINDS OF EFFECTIVENESS TEACHING?

Here the evidence is fairly clear. Not only can students distinguish between very good and very poor teachers, they can make finer distinctions as in the study by Marsh, Overall, and Kesler (1979) cited just above. Here the student evaluations significantly differentiated between two courses taught by the same instructor. The student ratings were higher on every evaluation factor for the course the teacher designated "more effective."

There is evidence from a large number of factor analytic studies that there are several underlying dimensions in student ratings. A typical breakdown of these dimensions would include a Skill scale (interesting presentation of material, intellectual stimulation, clarity). This is the most important dimension and relates most highly to items rating "all around teaching ability" or "all around value of the course." Other dimensions often found include Rapport (concern for students, classroom interaction), Structure (organization and preparation of the course), and Difficulty (amount of work demanded). (See Kulik and Kulik, 1974; Feldman, 1976b.) The appearance of these dimensions (or similar ones) in studies indicates that students do not simply make global "good-bad" evaluations, but tend to rate on several important scales.

CAN STUDENT RATINGS DIRECTLY SUPPLY INFORMATION THAT WILL ASSIST THE TEACHER IN IMPROVEMENT OF INSTRUCTION?

The record in this area is pretty disappointing. Kulik and McKeachie (1975) summarize the studies in this area by saying that there is no evidence that the information students provide helps teachers to improve their effectiveness. They point out some of the difficulties involved: (1) The studies typically look for changes in teacher effectiveness from mid-term to end-of-term. That is probably not enough time for significant changes to take place and be recognized by students. (2) Teachers in these studies are given feedback from their own students but no comparisons with other teachers. Without a set of norms for comparison they find it difficult to interpret their ratings. (3) Teachers may need assistance from instructional development experts to help effect changes. (The IDEA system addresses these difficulties by providing extensive norms and diagnostic assistance with the interpretation of results.) (See Erickson and Erickson, 1979; Hoyt and Howard, 1977, for descriptions of faculty development programs that use student ratings and the assistance of an instructional development expert.)

In a couple of studies, changes in teacher behavior following student ratings have been documented. Centra (1973) found that teachers who held unrealistic

tically high opinions of their teaching effectiveness at midsemester appeared to be most affected by feedback. In Pambookian's (1974) study, teachers who received average ratings changed more than teachers rated either favorably or unfavorably.

In all, student ratings taken alone, without assistance in interpretation, seem to be of limited help to a teacher wishing to improve instruction.

CAN STUDENT RATINGS BE COMBINED WITH AN EMPIRICAL RESEARCH BASE (CORRELATING STUDENT PROGRESS AND SPECIFIC TEACHER ACTIVITIES) TO SUPPLY INFORMATION THAT WILL ASSIST THE TEACHER IN IMPROVEMENT OF INSTRUCTION?

The IDEA system data pool now contains over 23,000 classes. From this pool a data base for all courses and for courses similar in motivation level and class size has been established for comparative purposes. Correlation tables relating specific student progress variables and specific teacher activities (called "methods") are stored in the computer. Feedback to teachers from IDEA includes not only the raw data from their students but the norming data from all classes and from the appropriate similar group of classes as well. IDEA provides a diagnostic summary in which appears a print-out of the teacher-selected progress variables (goals). For each of these goals in which progress is rated average or below by students, the teacher also receives a list of teacher activities which are correlated with the goal (and which were also related low by the students). This assistance in interpretation provides the teacher with a starting place for considering possible changes that may improve teaching effectiveness.

SUMMARY: In general, student rating forms are reliable instruments. The research in the area has identified certain powerful extraneous variables and it is possible to control for the most important of these. A case has been made in this paper for the validity of student ratings for evaluative purposes. We have established that student ratings are strongly related to the ratings of teacher effectiveness by administrators, peers, and (most important) the teachers themselves. We have also found that students can make some pretty fine distinctions along a number of dimensions of teacher effectiveness. It is important to note at this point that we are not suggesting that student ratings be the only measure of teacher effectiveness used in evaluations of teachers. We are suggesting that student ratings are adequately reliable and relevant measures to be included in teacher evaluation procedures.

We have also found that student ratings by themselves are not very helpful to teachers seeking to improve instruction. However, some student rating systems (notably IDEA) provide interpretive assistance which can enhance the usefulness of student ratings for diagnosis. It is also possible that instructional development specialists can provide help to teachers in translating the information from student ratings into appropriate changes in teacher behavior. Student ratings alone have limited validity for instructional development. Student ratings with interpretation and/or expert developmental assistance have far more promise.

* IDEA includes items describing teacher behaviors which are similar to those in most student rating forms. It has, in addition, several features which are useful in answering some of the typical objections to student ratings. For this reason IDEA will

be used in this paper as the primary example of a student rating instrument.

Footnote 1: Two types of reliability measures are regularly reported in the literature: internal consistency and stability. In reporting internal consistency the assumption is made that ratings of different students within a class are replicates of each other, any differences in their ratings being "error." (This is the most conservative assumption that can be made and would consistently underestimate reliability.) The reliability of a single rater is assumed to be equivalent to the agreement between two different students and is estimated by correlating random pairs of students. Typically these internal consistency measures are about .25 for the single rater and in the .80's and .90's for classes of twenty or more. (See Kulik and McKeachie, 1975; Feldman, 1977; Marsh and Overall, 1979.)

Stability coefficients (agreement in the ratings of the same student over time) are even higher. Here we assume that there is systematic variance associated with individual student characteristics (a sensible assumption as we shall see when we examine various factors which affect ratings). That is, not all variations among student ratings within a particular class are "error." Students do have "true" differences of opinion. Single rater measures of stability from the second to last hours of a course were reported by Kohlan (1973) at .58. Marsh and Overall (1979) found stability measures of .59 from end-of-course to a year after graduation. (Internal consistency measures in the same study were in the mid .20's.)

REFERENCES

- Bendig, A.W. Ability and personality characteristics of introductory psychology instructors rated competent and empathetic by the students. *Journal of Educational Research*, 1955, 48, 705-709.
- Blackburn, R. T. & Clark, M. J. An assessment of faculty performance: Some correlates between administrator, colleague, student, and self ratings. *Sociology of Education*, 1975, 48, 242-456.
- Centra, J. A. Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology*, 1973, 65, 395-401.
- Centra, J. A. Colleagues as raters of classroom instruction. *Journal of Higher Education*, 1975, 46, 327-337.
- Clegg, V. L. Teaching behaviors which stimulate student motivation to learn. Unpublished Doctoral Dissertation, Kansas State University, 1979.
- Costin, F., Greenough, W. T., & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 1971, 41, 511-535.
- Doyle, K. O., Jr. & Whitely, S. E. Student ratings as criteria for effective teaching. *American Educational Research Journal*, 1974, 11, 259-274.
- Elliot, D. N. Characteristics and relationships of various criteria of college and university teaching. *Purdue University Studies in Higher Education*, 1950, 70, 5-61.
- Erickson, G.R. & Erickson, B. L. Improving college teaching. *Journal of Higher Education*, 1979, 50, 670-683.
- Feldman, K. A. Grades and college students' evaluations of their courses and teachers. *Research in Higher Education*, 1976a, 4, 69-111.
- Feldman, K. A. The superior college teacher from the students' view. *Research in Higher Education*, 1976b, 5, 243-288.
- Feldman, K. A. Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education*, 1977, 6, 223-274.
- Feldman, K. A. Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 1978, 9, 199-242.
- Feldman, K. A. The significance of circumstances for college students' rating of their teachers and courses. *Research in Higher Education*, 1979, 10, 149-172.
- Frey, P. W. Student ratings of teaching: Validity of several rating factors. *Science*, 1973, 182, 83-85.
- Frey, P. W., Leonard, D. W., & Beatty, W. W. Student rating of instruction: Validation research. *American Educational Research Journal*, 1975, 12, 435-447.
- Frey, P. W. Validity of student instructional ratings: Does timing matter? *Journal of Higher Education*, 1976, 47, 327-336.
- Guthrie, E. R. The evaluation of teaching: A progress report. Seattle: University of Washington, 1954.
- Howard, G. S. & Schmeck, R. R. Relationship of changes in student motivation to student evaluations of instruction. *Research in Higher Education*, 1979, 10, 305-315.
- Hoyt, D. P. & Howard, G. S. The evaluation of faculty development programs. Manhattan, KS: Center for Faculty Evaluation and Development, 1977.
- Kohlan, R. G. A comparison of faculty evaluations early and late in the course. *Journal of Higher Education*, 1973, 44, 587-595.
- Kulik, J. A. & Kulik, C. C. Student ratings of instruction. *Teaching of Psychology*, 1974, 1, 51-57.
- Kulik, J. A. & McKeachie, W. J. The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), *Review of Research in Education*, Itasca, IL: F. E. Peacock, 1975.
- Lathrop, R. G. & Richmond, C. College students' evaluation of courses and instructors. Unpublished Manuscript, Chico State College, 1967.
- Lathrop, R. G. Unit factorial ratings by college student of courses and instructors. Unpublished Manuscript, Chico State College, 1968.
- Marsh, H. W. & Overall, J. V. Long-term stability of students' evaluations: A note on Feldman's "Consistency and variability among college students in rating their teachers and courses." *Research in Higher Education*, 1979, 10, 139-147.
- Marsh, H. W., Overall, J. V., & Kesler, S. P. Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluation and evaluations by their students.

REFERENCES

- Journal of Educational Psychology, 1979, 71, 149-160.
- Maslow, A. H. & Zimmerman, W. College teaching ability, scholarly activity and personality. Journal of Educational Psychology, 1956, 47, 185-189.
- Morsh, J. E., Burgess, G. G., & Smith, P. N. Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 1956, 47, 79-88.
- Pambookian, H. S. The initial level of student evaluation of instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology, 1974, 66, 52-56.
- Remmers, H. H. The relationship between students' marks and students' attitudes toward instructors. School and Society, 1928, 29, 759-760.
- Rodin, M. & Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1164-1166.
- Solomon, D., Rosenberg, L., & Bezdek, W. E. Teacher behavior and student learning. Journal of Educational Psychology, 1964, 55, 23-30.
- Sullivan, A. M. & Skanes, G. R. Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 1974, 66, 584-590.